Battle of Neighborhoods(Neighborhood of Schools in Manchester)

Coursera Capstone Final Project Report.

Introduction: Business Problem

- This project identifies and recommends the nearest schools and school categories in neighborhoods.

- People especially families may move from one settlement to the other without considering the status of schools and their categories within the city and its neighborhoods

Who will be Interested?

- The project targets families and students who will need to consider schools and their categories within the neighborhoods.

- The government will need to analyze schools and their categories within the neighborhoods to make the best choices for residents.

- Private sectors, investors and business institutions are also beneficiaries. These entities can secure the right locations to invest by making the right decision on which categories of schools will be of a high demand in various neighborhoods.

Data Requirement

• Based on the problem definition above, factors to be considered are:

 •Number of schools and their categories in a neighborhood

 •Radius distance of neighborhoods nearest to schools

• The following data sources will be required to extract relevant data:

 •A list of neighborhoods of Manchester from Wikipedia

 •A geocoder library for extracting latitude and longitude coordinates

 •FourSquare API for searching for schools within these neighborhoods and providing their various latitudes and longitude coordinates

Pre processed Data

	PostalCode	Town	Neighborhood	Latitude	Longitude
0	M1	Manchester	Piccadilly, City Centre, Market Street	53.479396	-2.231743
1	M2	Manchester	Deansgate, City Centre	53.474096	-2.251063
2	M4	Manchester	Ancoats, Northern Quarter, Strangeways	53.484998	-2.227603
3	M8	Manchester	Crumpsall, Cheetham Hill	53.520912	-2.241695
4	M9	Manchester	Harpurhey, Blackley	53.511907	-2.208449

Methodology

• First, the data has been retrieved and cleaned as explained previously. The data consists of schools within a radius of 1000m for each neighborhood with a limit set to 100.

• Next, we analyze the data by exploring the various categories to which the schools belong.

• After this, we explore the distribution in each neighborhood in a more explanatory way using maps for visualization. This gives us a good idea of the distribution of schools.

Methodology (Cont'd)

• Then for those neighborhoods without schools, we found the distance to the closest school by incrementing the radius by 100 meters in each iteration until a school is found using Folium.

• The number of schools in each neighborhood will be clustered using **K-Means**.

• Finally, we will discuss and conclude which neighborhoods require an improvement of facilities in terms of schools and which regions have very good facilities.
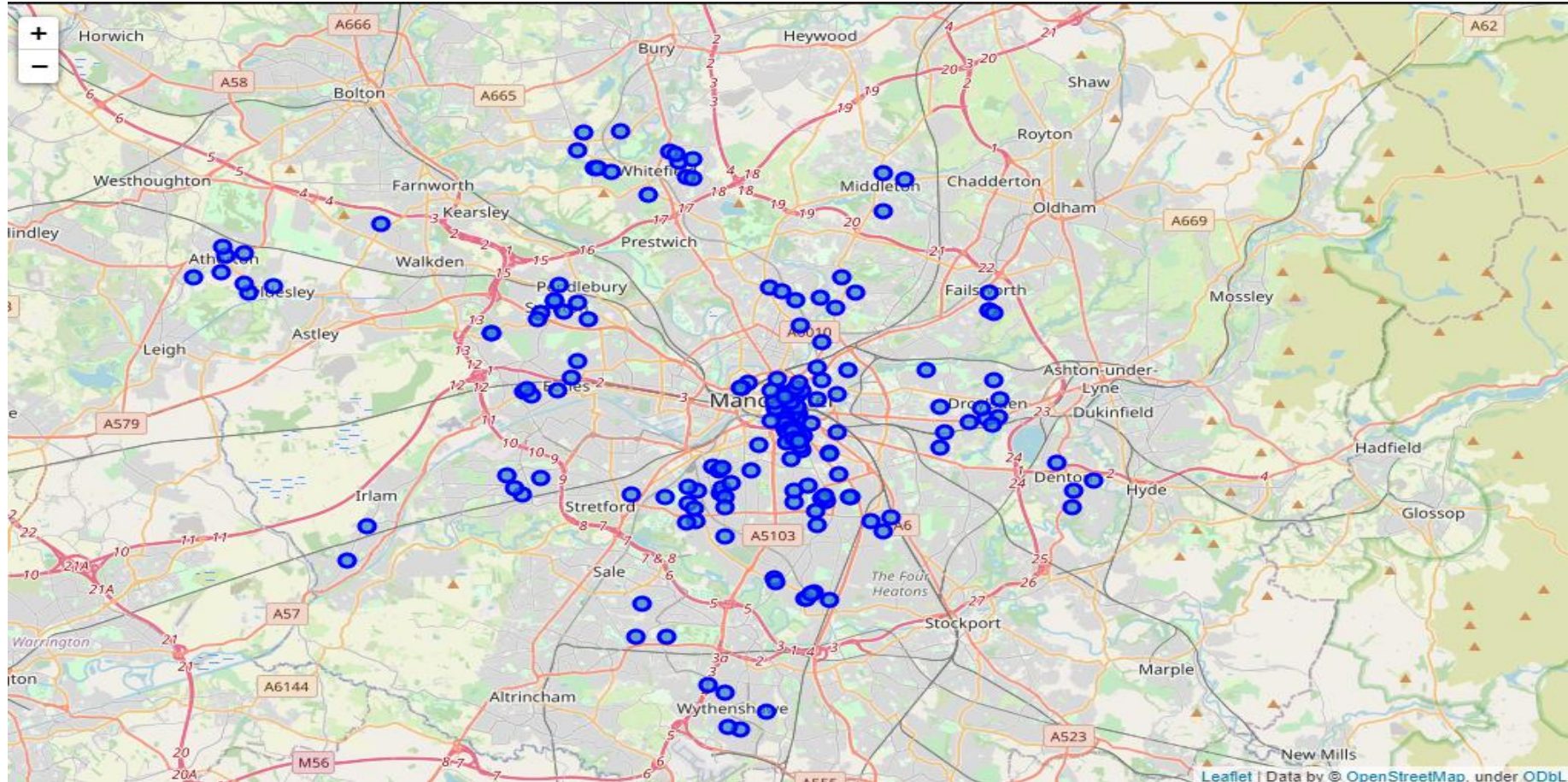
Analysis:

- Based on the data obtained from Folium about schools, I explored the categories to which schools belong.

- From the bar graph below, we observe that Elementary Schools are more than any other School Category.

- Law Schools, Medical Schools and Pre Schools are the least in number

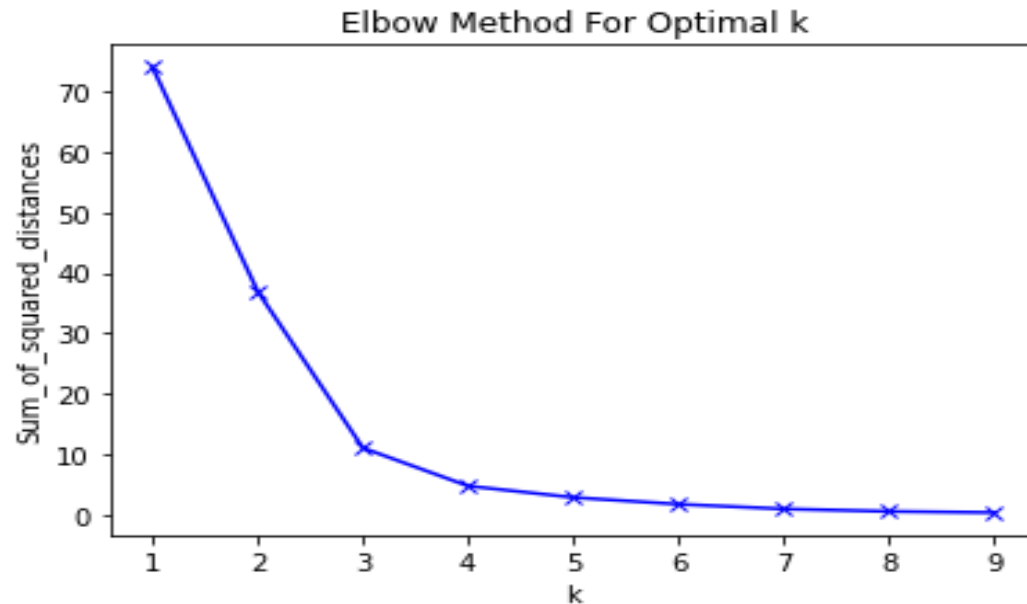# School Categories Bar Graph
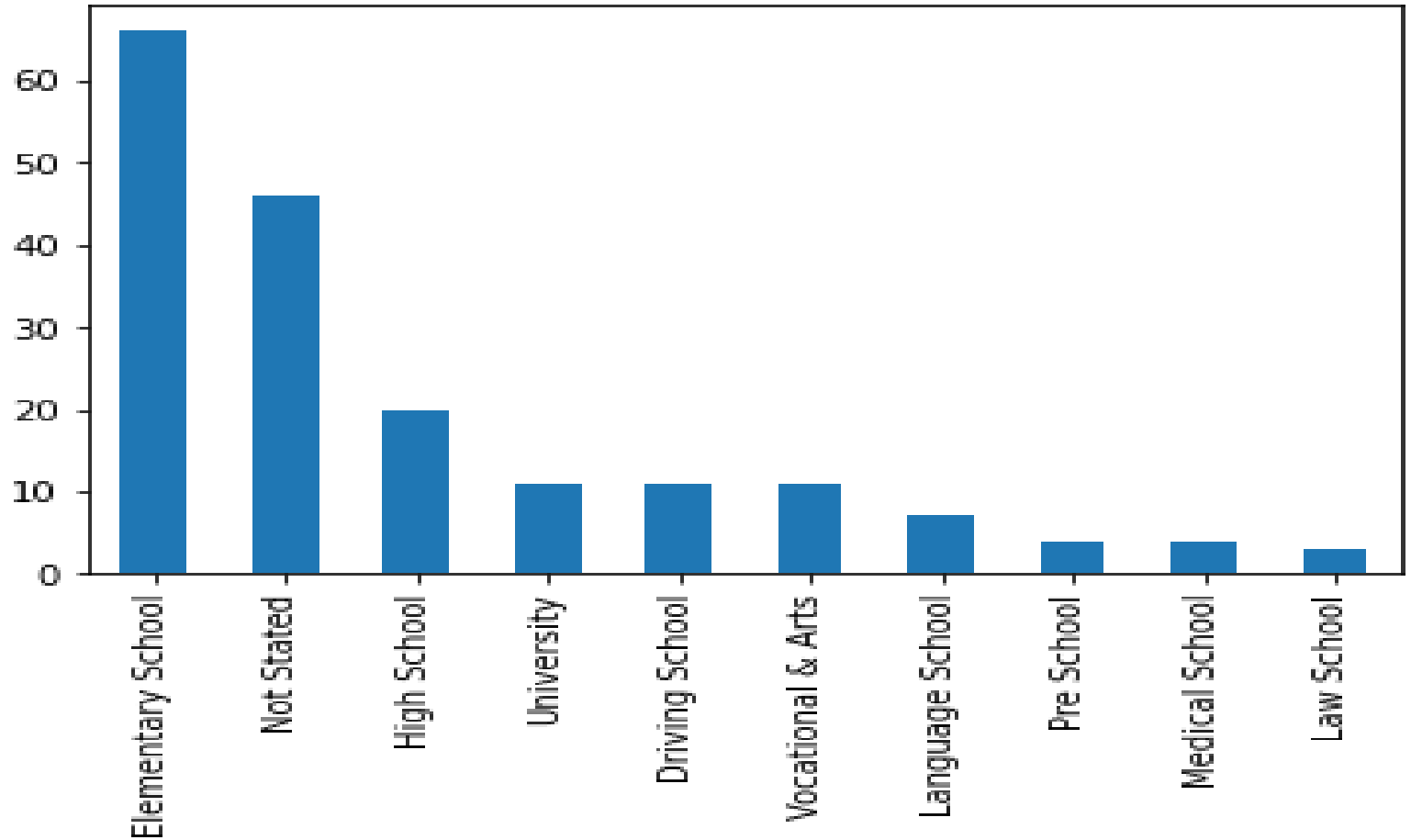
Distribution of Neighborhoods on Map

Clustering Neighborhoods

- We find the best cluster number to use by finding the mean of the radius of the data within the cluster number range of 1 to 10.

- Below is an image of the best number of cluster to use on our data.

Clustering (Cont'd)

- We will now cluster the neighborhoods using K-Means clustering algorithm into 3 clusters based on the distance to nearest school and number of schools in each neighborhood.

- The nearest distance to school is varied between a radius of 1000 meters to 2000 meters. The image below shows the neighborhoods with the nearest schools between 1000 meters and 2000 meters.

- These features form a good basis for identifying the neighborhoods requiring improvement.

Clustering (Cont'd)

- Neighborhoods without schools within 1000m radius. The first five records are displayed below:

```
    hoods_without_schools_df['Sch_in_1000m']=0
    hoods_without_schools_df.head()
```

	Neighborhood	Latitude	Longitude	Sch_Radius	Sch_in_1000m
0	Trafford Park, The Trafford Centre	53.471752	-2.320928	1600	0
1	Prestwich, Sedgley Park, Simister	53.530427	-2.296019	1100	0
2	Carrington, Partington	47.449720	-99.126224	1100	0
3	Manchester Airport	53.350342	-2.280369	1800	0

Clustering (Cont'd)

• After merging with neighborhoods with schools within the 1000 meters radius. The result showing just the first five(5) data:

	Neighborhood	Latitude	Longitude	Sch_Radius	Sch_in_1000m	label
0	Trafford Park, The Trafford Centre	53.471752	-2.320928	1600	0	1
1	Prestwich, Sedgley Park, Simister	53.530427	-2.296019	1100	0	0
2	Carrington, Partington	47.449720	-99.126224	1100	0	0
3	Manchester Airport	53.350342	-2.280369	1800	0	1
4	Abbey Hey, Gorton	53.469858	-2.152198	1000	6	0

Mean of Clustered Data

```
Merged_Hood_Radii_df.groupby('label').mean()
```

|	Latitude	Longitude	Sch_Radius	Sch_in_1000m
label				
0	53.303404	-5.121835	1005.882353	4.382353
1	53.411047	-2.300648	1700.000000	0.000000
2	53.479396	-2.231743	1000.000000	34.000000

Observation of Data Clustered

• As it is clear, the neighborhoods labeled 0 have schools **within 1000m** and a number of options in terms of schools with an average of **4.38**

• The neighborhoods labeled 1 are the ones **requiring improvement** as the average distance to nearest schools is close to **2km.**

• The neighborhoods labeled 2 also have easy access to good education facilities at school level within **1000m radius** with an average of about **34 schools** per neighborhood. Most of the neighborhoods fall under this category.

Final Map of Clusters in Manchester

- Cluster 1 in Blue, Cluster 2 in Green and Cluster 3 in Red
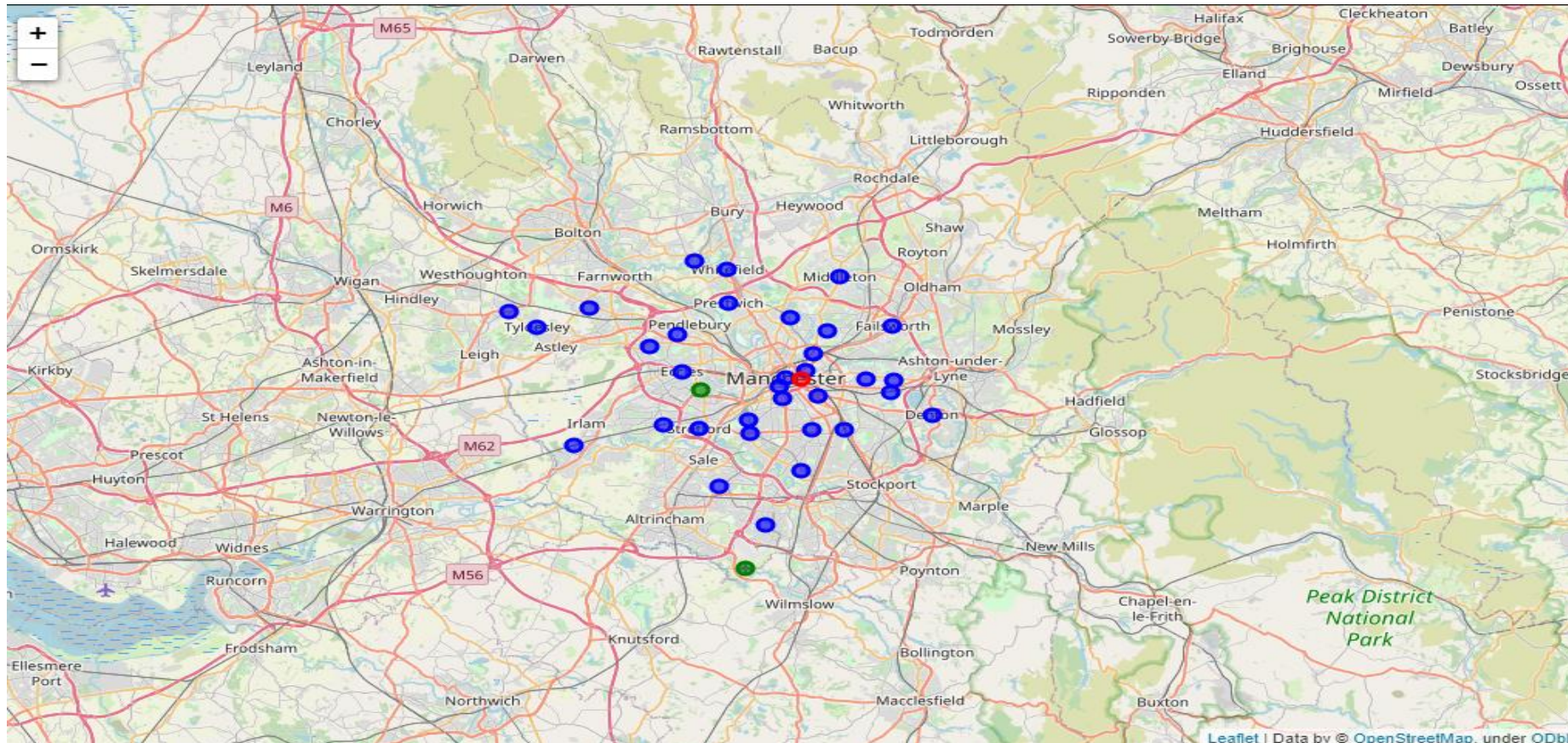
Results and Discussions

- we observe that schools providing education only at university level are comparatively less which is expected.

- we have observed the distribution of schools and find that mostly they are concentrated towards the center of the city and are less dense in the outskirts. This is clear from the map above.

- Finally, we have clustered the neighborhoods into 3 clusters and have observed that most of the neighborhoods fall under category with label 2.

Conclusion:

• In this analysis, the distribution of schools in Manchester has been analyzed. Also, the various neighborhoods have been clustered based on their access to school facilities per distance.

• The neighborhoods requiring improvement have been identified along with those neighborhoods which have the potential for profitable schools.

• Also, the various categories of schools have been observed.

• Link to Notebook Code: the battle of neighborhoods